Machine Learning Yearning

# End-to-end deep learning

饶正锋

Nov 2018

# outline

What is End-to-end?

Examples of End-to-end

End-to-end: Pros and cons

Pipeline or End-to-end ?

More

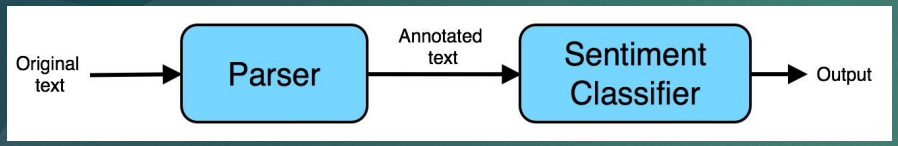
# What is End-to-end deep learning?

**[Task]** sentiment classification:

*This is a great mop!* --positive

*This mop is low quality--I regret buying it.* --negative

Pipeline(traditional ML):



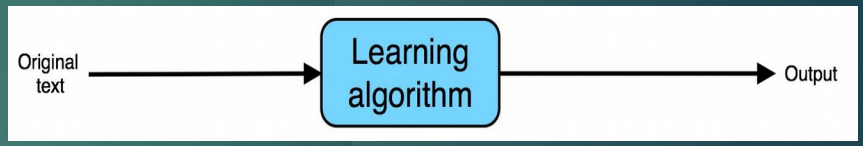

Two components:

1. parser: annotate texts, identify import words

This is a great$_{\text{Adjective}}$ mop$_{\text{Noun}}$!

2. sentiment classifier: class the annotated texts, giving adjectives a higher weight

End to end(Deep Learning):




Input the raw, original text:

*This is a great mop!*

Try to directly recognize the sentiment

**End-to-end:**

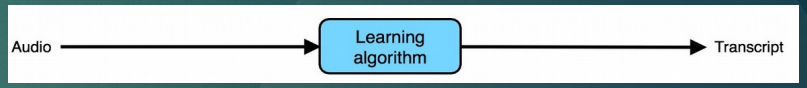
Asking the learning algorithm to go directly from the input to the desired output.

# Examples of End-to-end

**[Task]**speech recognition

Pipeline:


End to end:




Three components:

1. Compute features: Extract hand-designed features, such as MFCC

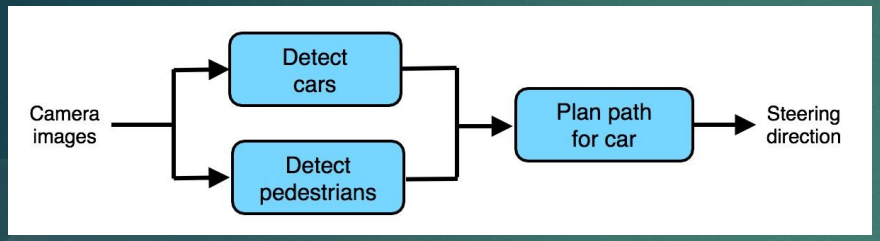
2. Phoneme recognizer:recognize the phonemes in the audio clip.

3.Final recognizer: Take the sequence of recognized phonemes, and try to string them together into an output transcript.

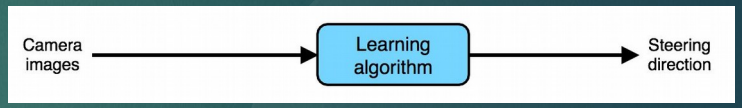
# Examples of End-to-end

**[Task]**autonomous car

Pipeline:




End to end:

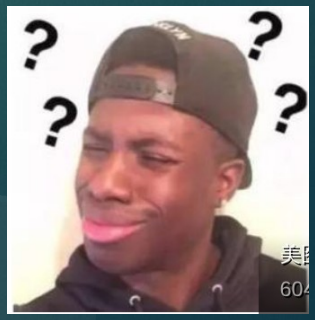



Three components:

1. Detect cars
2. Detect pedestrians
3. Plan path for car

Andrew Ng:

Even though end-to-end learning has seen many successes...but I'm skeptical about end-to-end learning for autonomous driving...



# End-to-end: Pros and cons

## Pros:

1. No hand-engineered needed

2. Won't be hampered by the limitations of hand-engineered features,some useful information might be dropped by hand-engineered features

3. Directly learning rich outputs

## Cons:

1. Needs **lots** of labeled data for "both ends"—the input end and the output end

**Labeled Data!!!**

Traditional supervised learning:
h: X→Y (y:integer / real number)

| Problem | X | Y |
|---|---|---|
| Spam classification | Email | Spam/Not spam (0/1) |
| Image recognition | Image | Integer label |
| Housing price prediction | Features of house | Price in dollars |
| Product recommendation | Product & user features | Chance of purchase |

End-to-end DL:

| Problem | X | Y | Example Citation |
|---|---|---|---|
| Image captioning | Image | Text | Mao et al., 2014 |
| Machine translation | English text | French text | Suskever et al., 2014 |
| Question answering | (Text,Question) pair | Answer text | Bordes et al., 2015 |
| Speech recognition | Audio | Transcription | Hannun et al., 2015 |
| TTS | Text features | Audio | van der Oord et al., 2016 |

when the training set is small, it might do worse than the hand-engineered pipeline

# End-to-end: Pros and cons

科技 中 双语 英

新新世界

**廉价劳动力如何推动中国的人工智能雄心**

袁莉

2018年11月26日

位于中国中部河南省郑县睿金科技公司总部的工人。 他们识别图像中的物体，以帮助人工智能理解世界。 YAN CONG FOR THE NEW YORK TIMES

向中国科技目标推进的一些最关键工

Show

在中国中部腹地的一座原水泥厂里

照片和视频，标记他们看到的每样东西。这是一辆汽车，那是一个红绿灯。这是面包，这是牛奶，那是巧克力。这是一个人走路的样子。

“我以前觉得机器很聪明，”24岁的侯夏梦说，“现在我知道，它所有的聪明都是我们给它的。”

在长期充当世界工厂的中国，新一代的低薪工人正在为未来奠定基础。在规模较小、成本较低的城市，创业公司如雨后春笋般涌现，他们在给中国海量的图像和监控录像添加标签。正如一位专家所说，如果把中国比做数据方面的沙特阿拉伯，那么这些公司就是炼油厂，它们将原始数据转化为可以为中国的人工智能雄心提供动力的燃料。

Show

可能会被浪费掉。目前还不清楚人工智能竞赛是否会是一场赢者通吃的零和博弈。除非有人能够分析和归类，否则数据是无用的。

但标记这些数据的能力，可能是中国真正的人工智能实力，也可能是美国唯一无法匹敌的力量。在中国，这个新兴产业提供了一个政府长期承诺的未来的一瞥：一个建立在科技而非制造业基础上的经济。

“我们这些人属于数据行业里的建筑工人。我们做的事就是垒砖头，一块一块地垒，”中部省份河南郏县一家数据标签工厂的联合创始人伊亚科说，“但我们在人工智能中扮演着重要角色，没有我们，他们无法建造摩天大楼。”

Show

由于商店照明和人体运动，图像变得更加复杂，系统无法分辨玛芬蛋糕、甜甜圈或叉烧包。AInnovation项目经理梁睿表示，标记在使用商店内部照片的情况下，准确度高达99%。

“有多少人工就有多少智能，”梁睿说。

AInnovation有不到30个标记员，但标签初创公司的激增使得把工作外包出去变得很容易。有一次，梁睿为超市做标签，需要在三天内拍摄大约20000张照片。同事们在数据工厂的帮助下，只花几千美元就完成了工作。

“我们就是10年前流水线上的工人，”河南那家数据工厂的联合创始人伊亚科说。

Show

# Pipeline or End-to-end ?

For specified task, choose pipeline or end-to-end, should consider:

1. **Data availability**

**[Task]** autonomous car

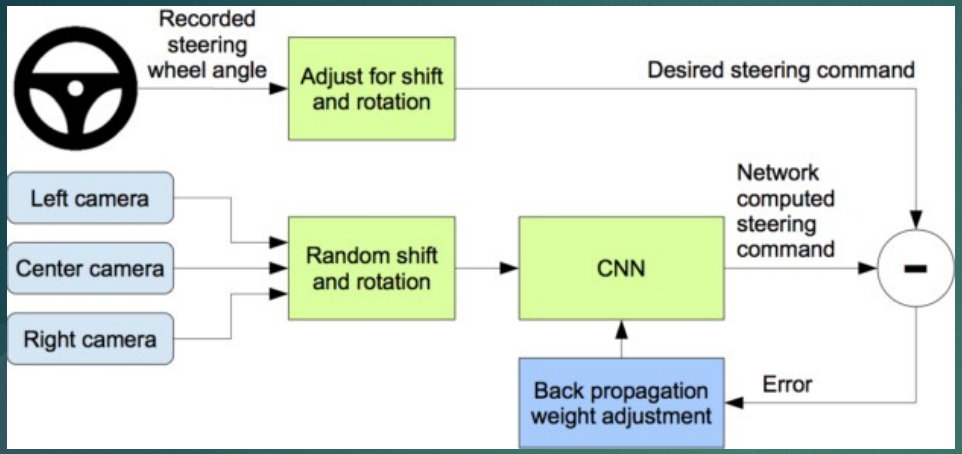
**End-to-end**:Need a large dataset of (Image, Steering Direction) pairs.
It is very time-consuming and expensive.

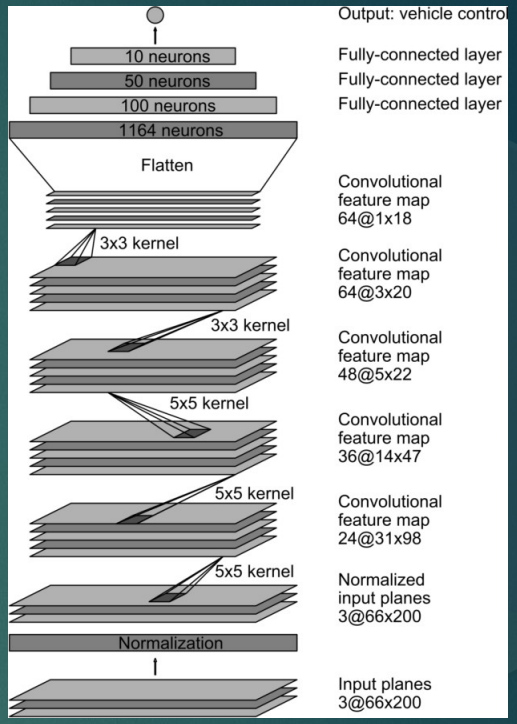
**Pipeline**: There are numerous computer vision datasets with large numbers of labeled cars and pedestrians, it's easy to build a car detector and a pedestrian detector with these data.

Really?

# Pipeline or End-to-end ?

End to End Learning for Self-Driving Cars(By Nvidia)




**Something interesting:**

1. Training data was collected by driving about 72 hours on a wide variety of roads and in a diverse set of lighting and weather conditions.
2. The training data is therefore augmented with additional images that show the car in different shifts from the center of the lane and rotations from the direction of the road.
3. It's able to learn the entire task of lane and road following without manual decomposition into road or lane marking detection, semantic abstraction, path planning, and control. A small amount of training data is needed.

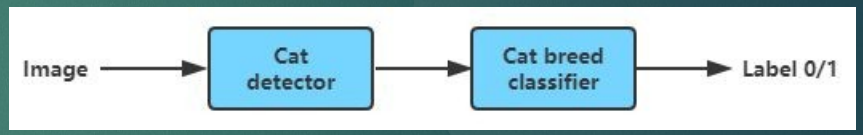
**It's the success of end-to-end deep learning!**

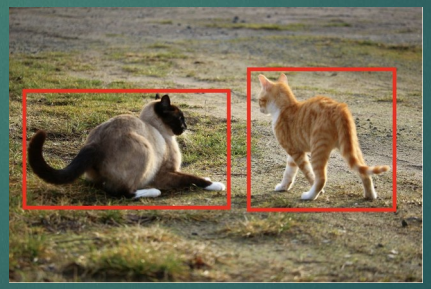
# Pipeline or End-to-end ?

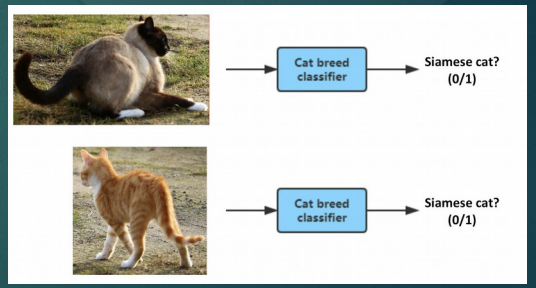
2. **Task simplicity**
**[Task]** Siamese cat detector.

End to end:


Pipeline:






Tips: how to define a task’s simplicity?

# More

东北话和川普，机器都能听懂，吴恩达说的端到端学习究竟是什么？
End to End Learning for Self-Driving Cars.pdf
End-to-End Deep Learning for Self-Driving Cars
端到端的TTS深度学习模型tacotron(中文语音合成)

# THANK YOU